Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A finite-state multimodal recognition system that generates a multimodal meaning based on an utterance comprising a plurality of associated modes, the system comprising:

means for receiving said utterance;

a plurality of finite-state mode recognition systems, each finite-state mode recognition system usable to recognize ones of the associated modes, each finite-state mode recognition system outputting at least one recognition lattice for each associated mode; and

an n-tape finite-state device that inputs n-1 recognition lattices from the plurality of finite-state mode recognition subsystems and outputs the multimodal meaning based on the n-1 recognition lattices.

2. (Currently Amended) A finite-state multimodal recognition system that generates a multimodal meaning based on an utterance comprising a pair of associated modes, the system comprising:

means for receiving said utterance;

a pair of finite-state mode recognition systems, each finite-state mode recognition system usable to recognize one of the associated modes, each finite-state mode recognition system outputting at least one recognition lattice for each associated mode; and

a multimodal recognition system that inputs a recognition lattice from each of the pair of mode recognition systems and outputs the multimodal meaning for the pair of associated modes based on the plurality of recognition results, comprising:

a first system that inputs the pair of recognition lattices and ~~that~~ outputs a combined recognition finite-state transducer;

a second system the inputs the combined recognition finite-state transducer and outputs a combined recognition finite-state machine, and

a third system that inputs the combined recognition finite-state machine and a multimodal meaning grammar and outputs the multimodal meaning.

3. (Currently Amended) A multimodal recognition system that generates a multimodal recognition based on an utterance comprising a plurality of associated modes, the system comprising:

means for receiving said utterance;

a plurality of mode recognition subsystems, each mode recognition subsystem usable to recognize ones of the associated modes, each mode recognition subsystem outputting at least one recognition result for each associated mode; and

a multimodal recognition subsystem that inputs recognition results from each of the plurality of mode recognition subsystems and outputs the multimodal recognition for the plurality of associated modes based on the plurality of recognition results;

wherein each of the plurality of mode recognition subsystems and the multimodal recognition subsystem includes at least one finite-state machine having at least one tape.

4. (Original) The multimodal recognition system of claim 3, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the recognition results from at least one of the plurality of mode recognition subsystems and that generates a first finite-state transducer that relates the input recognition results from each of the at least one mode recognition subsystems to a recognition model of at least one other mode recognition subsystem.

5. (Original) The multimodal recognition system of claim 4, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the recognition results from the at least one other mode recognition subsystem and that generates a second finite-state transducer based on the recognition results from the at least one other mode recognition subsystem and the first finite-state transducer.

6. (Currently Amended) The multimodal recognition system of claim 5, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a recognition ~~results relating~~ based on said at least one finite-state machine.

7. (Currently Amended) The multimodal recognition system of claim 6, wherein the multimodal recognition subsystem further comprises:

a third finite-state transducer; and

a multimodal recognizer that inputs ~~the first~~said at least one finite-state machine and outputs the multimodal recognition based on said at least one~~the first~~ finite-state machine and the third finite-state transducer.

8. (Original) The multimodal recognition system of claim 7, wherein the multimodal recognition is a multimodal meaning.

9. (Original) The multimodal recognition system of claim 4, wherein the first subsystem comprises:

at least one second finite-state transducer, each second finite-state transducer relating the recognition results of one of the plurality of mode recognition systems to the recognition model of the at least one other mode recognition subsystem; and

a second subsystem that generates the first finite-state transducer based on the input recognition results from the at least one mode recognition subsystem and the at least one second finite-state transducer.

10. (Original) The multimodal recognition system of claim 9, wherein the first subsystem further comprises a third subsystem that generates at least one projection of the first finite-state transducer, each projection output to a corresponding one of the at least one other mode recognition subsystem.

11. (Currently Amended) The multimodal recognition system of claim 10, wherein each projection output to a corresponding one of the at least one other mode recognition subsystem in said plurality is usable as a recognition model by ~~that~~ said at least one other mode recognition subsystem.

12. (Currently Amended) The multimodal recognition system of claim 10, wherein ~~each~~ said at least one other mode recognition subsystem inputs the corresponding projection as a recognition model usable to recognize the at least one associated mode that is recognized by ~~that~~ said at least one other mode recognition subsystem.

13. (Original) The multimodal recognition system of claim 3, wherein the plurality of mode recognition subsystems comprise at least two of a gesture recognition subsystem, a speech recognition subsystem, a pen input recognition subsystem, a computer vision recognition subsystem, a haptic recognition subsystem, a gaze recognition subsystem, and a body motion recognition system.

14. (Original) The multimodal recognition system of claim 13, wherein the plurality of mode recognition subsystems includes at least a first mode recognition subsystem that inputs a first one of the plurality of different modes and outputs a first mode recognition lattice as the recognition result of the first mode recognition subsystem and a second mode recognition subsystem that inputs a second one of the plurality of different modes and outputs a second mode recognition lattice as the recognition result of the second mode recognition subsystem.

15. (Original) The multimodal recognition system of claim 14, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the first mode recognition lattice from the first mode recognition subsystem and that generates a first finite-state transducer that relates the first mode recognition lattice to a recognition model of the second mode recognition subsystem.

16. (Original) The multimodal recognition system of claim 15, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the second mode recognition lattice from the second mode recognition subsystem and that generates a second finite-state transducer based on the second mode recognition lattice from the second mode recognition subsystem and the first finite-state transducer.

17. (Original) The multimodal recognition system of claim 16, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a first finite-state machine.

18. (Original) The multimodal recognition system of claim 17, wherein the multimodal recognition subsystem further comprises:

a third finite-state transducer; and

a multimodal recognizer that inputs the first finite-state machine and outputs the multimodal recognition based on the first finite-state machine and the third finite-state transducer.

19. (Currently Amended) The multimodal recognition system of claim 18, wherein:

the third finite-state transducer relates the first ~~mode~~ one of said plurality of modes and the second ~~mode~~ one of said plurality of modes to a meaning of a combination of said first one of said plurality of modes and said second one of said plurality of modes ~~the first and second modes~~; and

the multimodal recognizer comprises a meaning subsystem that inputs the first finite-state machine and outputs, as the multimodal recognition, a ~~possible~~ meaning lattice based on the first finite-state machine and the third finite-state transducer.

20. (Original) The multimodal recognition system of claim 15, wherein the first subsystem comprises:

a second finite-state transducer that relates the first mode recognition lattice from the first mode recognition system to the recognition model of the second mode recognition subsystem; and

a second subsystem that generates the first finite-state transducer based on the input first mode recognition lattice and the second finite-state transducer.

21. (Original) The multimodal recognition system of claim 20, wherein the first subsystem further comprises a third subsystem that generates a projection of the first finite-state transducer.

22. (Original) The multimodal recognition system of claim 21, wherein the projection is output to the second mode recognition subsystem and is usable as a recognition model by the second mode recognition subsystem.

23. (Original) The multimodal recognition system of claim 21, wherein the second mode recognition subsystem inputs the projection as a recognition model usable to recognize at least the second mode input by the second mode recognition subsystem.

24. (Original) The multimodal recognition system of claim 3, wherein the plurality of mode recognition subsystems includes at least a gesture recognition subsystem that inputs a gesture mode and outputs a gesture recognition lattice as the recognition result of the gesture recognition subsystem and a speech recognition subsystem that inputs at least one speech mode and outputs a word sequences lattice as the recognition result of the speech recognition subsystem.

25. (Original) The multimodal recognition system of claim 24, wherein the multimodal recognition subsystem comprises a first subsystem that inputs the gesture recognition lattice from the gesture recognition subsystem and that generates a first finite-state transducer that relates the gesture recognition lattice to a recognition model of the speech recognition subsystem.

26. (Original) The multimodal recognition system of claim 23, wherein the multimodal recognition subsystem further comprises a second subsystem that inputs the first finite-state transducer and the word sequences lattice from the speech recognition subsystem and that generates a second finite-state transducer based on the word sequences lattice from the speech recognition subsystem and the first finite-state transducer.

27. (Original) The multimodal recognition system of claim 26, wherein the multimodal recognition subsystem further comprises a third subsystem that inputs the second finite-state transducer and outputs a first finite-state machine.

28. (Original) The multimodal recognition system of claim 27, wherein the multimodal recognition subsystem further comprises:

a third finite-state transducer; and

a multimodal recognizer that inputs the first finite-state machine and outputs the multimodal recognition based on the first finite-state machine and the third finite-state transducer.

29. (Currently Amended) The multimodal recognition system of claim 28, wherein:

the third finite-state transducer relates ~~the~~ a gesture mode and ~~the~~ a speech mode to a meaning of a combination of the gesture and speech modes; and

the multimodal recognizer comprises a meaning subsystem that inputs the first finite-state machine and outputs, as the multimodal recognition, a ~~possible~~ meaning lattice based on the first finite-state machine and the third finite-state transducer.

30. (Original) The multimodal recognition system of claim 25, wherein the first subsystem comprises:

a second transducer that relates the gesture recognition lattice from the gesture recognition systems to the recognition model of the speech recognition subsystem; and

a second subsystem that generates the first finite-state transducer based on the input gesture recognition lattice and the second finite-state transducer.

31. (Original) The multimodal recognition system of claim 30, wherein the first subsystem further comprises a third subsystem that generates a projection of the first finite-state transducer.

32. (Original) The multimodal recognition system of claim 31, wherein the projection is output to the speech recognition subsystem and is usable as a recognition model by the speech recognition subsystem.

33. (Original) The multimodal recognition system of claim 31, wherein the speech recognition subsystem inputs the projection as a recognition model usable to recognize the at least one speech mode input by the speech recognition subsystem.

34. (Original) The multimodal recognition system of claim 25, wherein the first subsystem comprises:

a second finite-state transducer that relates the gesture recognition lattice from the gesture recognition system to a language model of the speech recognition system as the recognition model of the speech recognition subsystem; and

a second subsystem that generates, as the first finite-state transducer, a gesture/language model finite-state transducer based on the input gesture recognition lattice and the second finite-state transducer.

35. (Original) The multimodal recognition system of claim 34, wherein the first subsystem further comprises a third subsystem that generates a projection of the gesture/language model finite-state transducer.

36. (Original) The multimodal recognition system of claim 35, wherein the projection is output to the speech recognition subsystem and is usable as a language model by the speech recognition subsystem.

37. The multimodal recognition system of claim 35, wherein the speech recognition subsystem inputs the projection as a language model usable to recognize the at least one speech mode input by the speech recognition subsystem.

38. (Original) The multimodal recognition system of claim 25, wherein the recognition model is one of a grammar model or a language model.

39. (Original) The multimodal recognition system of claim 24, wherein the gesture recognition subsystem comprises a gesture feature extraction subsystem that inputs the gesture mode and outputs a gesture feature lattice and a gesture recognition subsystem that inputs the gesture feature lattice and outputs the gesture recognition lattice.

40. (Original) The multimodal recognition system of claim 24, wherein the speech recognition system comprises:

a speech processing subsystem that inputs a speech signal and outputs a feature vector lattice;

a phonetic recognition subsystem that inputs the feature vector lattice and an acoustic model lattice and outputs a phone lattice;

a word recognition subsystem that inputs the phone lattice and a lexicon lattice and outputs a word lattice; and

a speech mode recognition subsystem that inputs the word lattice and a recognition model and outputs the word sequences lattice.

41. (Original) The multimodal recognition system of claim 40, wherein the recognition model is input from the multimodal recognition subsystem.

42. (Original) The multimodal recognition system of claim 3, further comprising a plurality of mode input devices, at least two if the plurality of mode input devices inputting different modes.

43. (Original) The multimodal recognition system of claim 42, wherein the plurality of mode input devices comprises at least two of a gesture input device, a speech input device, a pen input device, a computer vision device, a haptic input device, a gaze input device, and a body motion input device.

44. (Original) The multimodal recognition system of claim 43, wherein at least two of the plurality of input devices are combined into a single multimodal input device.

45. (Currently Amended) A method for recognizing a multimodal utterance comprising a plurality of different modes, the method comprising:

receiving said multimodal utterance;

inputting at least a first mode of the multimodal utterance and a second mode of the multimodal utterance that is different than the first mode;

generating a first mode recognition lattice from the first mode;

generating a second mode recognition lattice from the second mode;

generating a first finite-state transducer based on the first mode recognition lattice and a second finite-state transducer;

generating a third finite-state transducer based on the first finite-state transducer and the second mode recognition lattice;

converting the third finite-state transducer to a first finite-state machine; and

generating a multimodal recognition based on the first finite-state machine and a fourth finite-state transducer.

46. (Original) The method of claim 45, wherein:

the fourth finite-state transducer relates the first mode and the second mode to a meaning of a combination of the first and second modes; and

generating the multimodal recognition based on the first finite-state machine and the fourth finite-state transducer comprises generating a possible meaning lattice based on the first finite-state machine and the fourth finite-state transducer.

47. (Original) The method of claim 46, wherein generating the first finite-state transducer based on the first mode recognition lattice and the second finite-state transducer comprises generating the first finite-state transducer based on the input first mode recognition lattice and a first mode-to-second mode finite-state transducer.

48. (Original) The method of claim 45, further comprising:

generating a projection of the first finite-state transducer; and

outputting the projection to the second mode recognition subsystem, wherein the projection is usable as a recognition model by the second mode recognition subsystem.

49. (Original) The method of claim 48, wherein generating the second mode recognition lattice from the second mode comprises recognizing the second mode using the projection as the recognition model usable in recognizing the second mode.

50. (Original) The method of claim 45, wherein generating the first mode recognition lattice from the first mode comprises:

extracting a plurality of first mode features from the first mode; and

generating the first mode recognition lattice from the extracted features.

51. (Original) The method of claim 45, wherein the first mode is one of a gesture mode, a speech mode, a pen input mode, a computer vision input mode, a haptic mode, a gaze mode, and a body motion mode.

52. (Original) The method of claim 51, wherein the second mode is a different one of the gesture mode, the speech mode, the pen input mode, the computer vision mode, the haptic mode, the gaze mode, and the body motion mode.

53. (Original) The method of claim 45, wherein the first mode is a gesture mode and the second mode is a speech mode.